The Psychoacoustics of Unease: A Comparison Between Granular and Physical Modeling Synthesis

Ariel Yixuan Liu

Lord Byng Secondary School, Vancouver, Canada
arielliu0406@gmail.com

Abstract. This paper investigates the psychoacoustic responses of unease produced by two popular techniques of sound synthesis in the modern age: granular synthesis and physical modeling synthesis. Both methods have the capability of generating unsettling sounds, which are discussed and analyzed based on their techniques and applications. Granular synthesis evokes discomfort through randomness, overlapping grains, and roughness. In contrast, physical modeling synthesis unsettles listeners through hyper-realistic instrument simulations, triggering the "uncanny valley" of sound. Comparing these two techniques, this study highlights the unique ways in which each synthesis method influences human auditory systems and findings suggest that physical modeling synthesis may provoke a strong sense of unease, offering valuable implications for sound design choices in music in the future.

Keywords: Computational Sound Synthesis, Granular Synthesis, Physical Modelling, Psychoacoustics.

1. Introduction

Psychoacoustics is the study of how humans perceive sound, focusing on auditory responses to surrounding stimuli and how they may affect the human mind to simulate various emotions. It examines the relationship between physical properties of sound and the perceived sensations of listeners [1]. Understanding the principles of psychoacoustics is critical for successful applications in a wide range of domains such as music synthesizers [2], sound therapy, and interactive media design [3]. By exploring the profound connection between sound and emotion, people can design experiences that better fit their audiences to create a more effective and captivating experience overall. One particular area where psychoacoustics is seen is in sound synthesis: the electronic process of producing sounds with no acoustic source which generates sound waves that can be applied in various ways [4]. This computational method of producing electrical signals has paved the way for various sound synthesis techniques, including the distinct approaches of granular synthesis, and physical modeling [2]. Research on the two synthesis techniques has a long history, with granular synthesis first proposed by Iannis Xenakis in 1971 [5], and early experiments with physical modelling conducted in the 1960s by researchers Kelly, Lochbaum, Ruis, and Hiller [6].

Granular synthesis is a well-known audio generation technique that involves rearranging tiny snippets of waveforms that range from 10-100ms known as "grains". These grains are analyzed using acoustic descriptors to create a map of their similarities, which allows the composer or system to control how the grains are used and manipulated in the final sound [7]. This method is frequently seen used in electronic music, where musicians will utilize granular effects in modern films and video game scores, enhancing and expanding their genres [8].

In contrast, physical modelling takes real world systems, such as acoustic instruments, and creates a model based on a set of mathematical equations to describe their behavior. Implementing the physical models involves creating a "discrete-time simulation algorithm," where the computer calculates the sound at each specific sample and eventually connects them together to become a continuous line of sound. In this way, physical modeling produces a realistic sound that can be controlled and manipulated through quantitative parameters [6].

Both techniques have been extensively studied, with research exploring various aspects of their technical implementation, algorithmic development, and the many sound design possibilities that are available. However, there is limited information about the psychoacoustic responses they evoke in listeners, and how these generated sounds have the potential to elicit emotions in different contexts. The gap in understanding the emotional and perceptual effects of granular and physical modeling synthesis hinders the development of essential modern applications in music, where soundscapes and immersive audio design play a critical role in shaping listener experience and auditory comfort. Specifically, the "Uncanny Valley" of sound, where specific frequencies or pitches of dissonance can stir negative emotions of fear or unease within one's mind [9].

This study examines how granular and physical modeling synthesis techniques influence significant psychoacoustic responses, particularly associated with the feelings of unease and fear, and explores how sounds can be intentionally or unintentionally designed to provoke these sensations. The aim is to acknowledge the significant effects that granular and physical modelling synthesis possess to provide valuable insight into how musicians, sound designers, and developers can interact with these two specific synthesis techniques to manipulate auditory responses. To achieve this, the paper will first dive deeper into the specific concepts of psychoacoustics relating to the feeling of unease and tension; next, granular synthesis is explored along with a more detailed literature analysis into its specific mechanisms and documented psychoacoustic effects; this is then followed by a clear review and analysis of physical modeling synthesis and its reported responses from previous research; finally, the paper will compare the two methods discussed and observe trends and highlight key takeaways for future sound design and directions.

2. Psychoacoustics of unease

2.1. Sources of roughness

When two tones have frequencies that differentiate by less than 20 Hz, the listener may perceive pulse-like fluctuations in loudness. As the frequency separation increases, subsequent amplitude fluctuations occurring within the range of 20-200 Hz contribute to a particular sound attribute commonly known as "sensory dissonance" or "roughness" [10]. Without a clean resolution between the two tones, the ear and brain receive an impure or harsh sound that may trigger unpleasant feelings and emotions that bring an unnatural or uncanny feeling because of its atonal nature. The brain often processes this as mismatch negativity (MMN) [10]. Additionally, other sources of roughness exist in various forms. A certain aspect such as timbre has the possibility to create roughness even without defined intervals or pitches. Timbre can be defined as the perceptual quality or "color" of a sound, with different instruments all possessing different characteristics that distinguish one from another. What makes timbre a possible source of roughness can come from the instrument's spectral shape, amplitude envelopes, or harmonic content [11]. Most notably, irregular spectral shapes retain sparse harmonics and envelopes, creating uneven distributions of energy across frequencies, and producing a sound that interferes with the brain's ability to perceive them as coherent sounds. This attribute possesses such power that a small change alone can shift the perceived emotion [11].

2.2. Biological and cultural exposure

These aspects may be closely tied to biological and cultural exposures, despite their apparent universality. In many cases, physical properties of sound and human auditory processing come as an automatic response, yet cultural influence and music training can modify the conscious judgment toward the sounds [12]. Previous studies have shown that through cultural evolution of musical genres, musical distortion and noise manipulation reproduces a sort of nonlinear phenomena (NLP) seen in animal calls [13]. These sounds are produced under "controlled roughness" to generate a rough yet musically appealing tone which is accepted among specific culturally and artistically harnessed groups [13]. A prominent example of this controlled roughness can be seen in Asian originated throat singing or overtone singing technique, where harmonics are produced from glottal closure; as the high and low notes are produced at the same time, a low buzzing or a constant drone can be heard, acting as a source of roughness that brings an eerie feeling to the listener [14]. Another closely linked example is the Lithuanian Sutartinės singing, characterized by beat diaphony, where closely spaced intervals are intentionally utilized to produce roughness [12]. As a result, the auditory perception of unease can be characterized through the various examples of roughness along with combinations of different spectral irregularities, timbres, and frequencies. Understanding how these attributes contribute to perceived sound provides a solid foundation to exploring deeper into granular and physical modelling synthesis and how these synthesis techniques can reproduce or exaggerate these features to evoke similar emotional and psychoacoustic responses.

3. Granular synthesis and perceptual responses

A single sine wave could become something disorienting. When analyzing the technique of granular synthesis, the concept of "overlapping" plays a significant role. As discussed before, granular synthesis involves slicing audio samples and layering them according to time, pitch, and space. Overlapping these grains causes sounds to interact in different ways. Most notably, superimpositions create complex textures through the randomness imposed on the grains [5]. By scattering and stretching overlapped micro-grains, granular synthesis allows for sound manipulation beyond smooth textures by utilizing jitter settings and modifying or randomizing parameters such as grain start time, grain density, attack and decay, pitch, amplitude, and inter-onset times [15]. The result varies from smooth stable textures to unsettling buzzing, or even ethereal sci-fi sound effects used in electronic music or movies [8]. When these grains intertwine, these once insignificantly short snippets of sounds could potentially generate buzzing and fluctuations previously mentioned in the sources of roughness.

3.1. Synchronized granular synthesis

When grains are spaced equally using a set variable, a linear temporal relationship is established. Each grain begins at a predictable point in time, producing stable and coherent texture [15].

$$x(t) = \sum_{n=0}^{N} g_n(t - t_n) \cdot w_n(t - t_n) \tag{1}$$

When analyzing this general formula for granular synthesis, it is visible that final output sound is dependent on several key variables: each grain's waveform (g_n), the applied envelope (w_n), and the starting time of each grain (t_n). During the process of synchronous granular synthesis, $t_n$is kept constant to create the fixed timing and stable texture. From a psychoacoustic standpoint, this more controlled manipulation technique of grain timing allows for less irregularities and unpredictability, resulting in a less unsettling sound compared to asynchronous synthesis or randomized granular textures. Additionally, experimenting with grain density will also affect the timbre of the sound. Low densities, usually under 30 grains per second, produce rhythmic effects, unlike higher density sounds, blurring rhythm and creating one stable pitch. As density increases, a richer timbre can be noticed due to the additional frequency components and modulations [15]. With the application of higher grain density, the human perception of the sound will be relatively pleasant and less unsettling, because of the controlled manipulation of timing and overlapping, compared to sparce grain density.

3.2. Asynchronous granular synthesis

Alternatively, asynchronous granular synthesis involves adding grains with no strict linear relationship. This introduces randomized elements to distribute grains across a space without concern to timing, producing an indistinct connection between density and placement and generating a rather unpleasant or unpredictable sound [15]. Similarly, in synchronous granular synthesis, grain density, amplitude, pitch, and other parameters are likewise applied to asynchronous granular synthesis. A significant difference between the two techniques lies in the parameter of inter-onset intervals (IOI), which in this case, is constantly changing to produce the inconsistent timing changes. With the constant varying of IOIs, overlapping of grains becomes less controlled, pitch perception is now difficult to identify, and textures blend in a way that produces unnecessary roughness, often sounding unpleasant or alien to the ear.

3.3. Spatial granular synthesis

Beyond grain density, amplitude, and IOIs, spatial arrangement plays a role in how listeners may feel. Listener Envelopment (LEV) is used to describe how immersed a listener feels by a certain sound, reflecting their perception of envelopment. In spatial granular synthesis, grains of sound are controlled in a way that affects where and when those sound events happen, evoking different feelings based on spatial awareness. Past research evidence points to the conclusion that shorter time intervals, around 20 milliseconds or less, with the support of even spread across space, create sounds that surround the listener known as diffuse envelopment. This can be seen as stable and smooth attributes also noticeable in synchronous granular synthesis. However, longer intervals or uneven distribution of sound will create irregular distributions of spread, producing more separate or localized sounds [16]. This evidently connects to the characteristics of asynchronous grains, where irregularity affects LEV and evokes discomfort contributing to the feeling of unease. Ultimately, the key concept still lies in the hands of distribution, overlap [9], and randomness that will amplify the effects on psychological responses.

4. Physical Modeling Synthesis and perceptual responses

Physical modeling synthesis (PMS) combines properties of real-world instruments and sound sources with mathematical models to simulate the physical characteristics of those sounds. In the real world, for example, an acoustic string instrument: each pluck varies in frequency, waves, and vibrations. These variables, along with dynamics and articulations, are constantly shifting, giving it the realism factor of live music. However, in physical modeling synthesis, every time a program or algorithm is run, the sound of the modeled instrument remains the same. Each run, the pitch, vibrations, oscillations, and waves are all programmed in a robotic manner [17]. The aspects of real instruments and the deviations in tone and various variables will never be fully and perfectly achieved and replicated by the models, which evokes the "uncanny valley" feel to the sound [9].

4.1. Components of physical modeling

There are two main components that PMS relies on: the exciter and the resonator. These components can then be modeled to analyze their connection and their impact on the psychoacoustic responses to PMS.

The exciter is used to strike or introduce initial energy to an object to generate vibrations. It is usually dependent on a variety of factors including the type of action, whether it's simulating the plucking of a string, or striking of a key; force and intensity, affecting both amplitude and dynamics; point of contact, changing timbre and tone quality; and the speed of excitation, adding articulation and texture to the sound [18]. These factors are all parts of the action that leads to a sound, simulating the way energy is applied to the instrument and determining the characteristics of the resulting vibrations in the chosen instrument model.

The resonator represents the response of the modeled instrument. It acts because of the exciter's action, and shapes the sound's decay, timbre, and tone based on the characteristics of the instrument. Vibrations are modeled, capturing the way size and shape may affect the final sound produced. In the modern world of physical modeling synthesis, resonators are often represented as a linear and time-invariant system, describing it in terms of modes of vibration [6]. Each mode has a specific shape, frequency, and damping factor. In synthesis, these modes can be combined similarly to additive synthesis, but with physically determined constants.

4.2. Applications for Physical Modeling Synthesis

From speech synthesis to musical instruments, PMS has been applied in various settings, demonstrating its impressive abilities to mimic real instruments at a cost of perceptual challenges that evoke the "uncanny valley" of sound.

One of the earliest applications of PMS was through speech synthesis. The Voder (1939), developed by Homer Dudley, was one of the first electronic speech synthesizers [17]. Although it was able to produce impressive pronunciations and tones, some sounds were rather monotonous and robotic. Lacking subtle microvariations, the generated speech audio came with a psychoacoustic feeling of unease. This early example highlights a key principle that even when accurately simulates the mechanics of sound production, even the tiniest deviations in variables can create a perceptual sense of tension or unease.

Taking another example of the violin, variables such as string tension, bow speed, and resonance can be manipulated to alter the sound of the models. When micro fluctuations in vibrations occur in real life, the strings can capture them in a way that amplifies their realism. However, when attempting to program a model that can reflect the exact fluctuations of a real violin, algorithms may be able to resemble the original but may also produce a timbre that sounds in some sort of way "off" or unsettling [19]. Micro-inharmonicities or beating, as mentioned in the sources of roughness, can also affect pitch and tone, distort sound and create the unpleasant roughness that evokes negative feelings.

5. Comparing granular and Physical Modeling Synthesis

After analyzing the two very different styles of sound synthesis techniques, it is evident that both methods have the capability to evoke a sense of unease and uncanniness, despite their different systems. Comparing the two techniques, granular synthesis acts as a rather more old-school method, with technical variables that can be manipulated within its grains. Physical modeling synthesis utilizes a more systematic and algorithmic approach, involving greater mathematical consideration as well as theoretical modeling of real-world physics. Both techniques are popular, appearing in many movies, electronic music, sound effects, and video games, and have been changing the sound of music since the 1900s.

Now, comparing granular and physical modeling synthesis from a more psychoacoustic perspective, two main differences can be noticed. While both methods have the ability to elicit unease, it is clear that they utilize different ways of provoking those feelings.

Granular synthesis, through the power of overlapping and randomness, creates such feelings via audible and highly noticeable roughness. This is influenced by the many variables that can be experimented with, causing pitches and frequencies to clash and produce unpleasant harmonics. The beating and fluctuations are also a highly prominent feature of granular synthesis, making it particularly effective in evoking these feelings.

Physical modeling synthesis instead uses the power of "real yet not too real" sounds to evoke unease, rather than unpredictable roughness or frequency discomforts. The sounds created by PMS are often slightly artificial and inhuman, triggering an unconscious sense of a fake or alien tone. These modeled sounds, which represent instruments, create a sense of hyper-realism at the cost of perceptual errors that trigger the uncanny valley effect in audio.

6. Summary

Now, stepping back to view the full picture, the impact of sound synthesis lies not only in the technical designs of the synthesis techniques, but also in how listeners perceive and react to the produced sounds. At the heart of this research, conclusions can be drawn from the two synthesis techniques. Both granular synthesis and physical modeling synthesis can generate new sounds and evoking distinct psychoacoustic feelings of unease, unpleasantness, or uncanniness, three qualities that contribute to negative and discomforting sounds often found in genres like horror and sci-fi, intentional or unintentional. Yet these two techniques differ in how they contribute to these feelings: granular synthesis elicits them through unpredictable randomness and the fluctuating beating of overlapping grains that cause roughness, while physical modeling synthesis provokes them through the hyper-realism of physical models and the micro-errors in algorithmic parameters when imitating real instruments, which generate uncanny, unreal qualities.

However, these points converge on a critical question: which method proves more effective in evoking unease? Based on this study, it can be concluded that physical modeling synthesis, through its blend of realism and perceptual errors, more strongly triggers the uncanny valley effect and thus has a heightened capacity to evoke unsettling responses in listeners. Granular synthesis, while also capable of producing such feelings, tends to rely on conscious and noticeable overlaps or glitch-like textures, whereas physical modeling synthesis unsettles the unconscious mind, provoking a deeper instinctual response tied to the fight-or-flight mechanism. Therefore, it is not noise or chaos that frightens the human mind, but the sliver of a boundary of real and artificial that proves the most disturbing and unsettling sounds to the human ear.

With these conclusions drawn, future research may explore how these psychoacoustic responses can be used strategically in fields such as interactive media, film scoring, and virtual reality, where unsettling sound plays a significant role in shaping audience experiences.

References

[1] "Psychoacoustics | Acoustic Research, " Acoustic Research, 2018. https://www.acousticresearch.org/psychoacoustics (accessed Sep. 12, 2025).

[2] "Sound Synthesis, " Gsu.edu, 2025. http: //hyperphysics.phy-astr.gsu.edu/hbase/Audio/synth.html (accessed Sep. 13, 2025).

[3] F. Sampayo, "Psychoacoustics: Decoding Sound's Impact on Mind and Health, " Schallertech, Feb. 03, 2024. https: //schallertech.com/en/psychoacoustics-the-effect-of-sound-on-human-perception/

[4] "Sound_Synthesis, " www.sfu.ca. https: //www.sfu.ca/sonic-studio-webdav/handbook/Sound_Synthesis.html (accessed Sep. 13, 2025).

[5] Simon Fraser University, "Granular Synthesis, " Www.sfu.ca, 2020. https: //www.sfu.ca/~truax/gran.html

[6] S. Bilbao et al., "Physical Modeling, Algorithms, and Sound Synthesis: The NESS Project, " Computer Music Journal, vol. 43, no. 2–3, pp. 15–30, 2019, doi: https: //doi.org/10.1162/comj_a_00516.

[7] A. Bitton, Philippe Esling, and T. Harada, "Neural Granular Sound Synthesis, " arXiv (Cornell University), Jan. 2020, doi: https: //doi.org/10.48550/arxiv.2008.01393.

[8] Native Instruments, "Granular synthesis: a beginner's guide, " Native Instruments Blog, Apr. 04, 2023. https: //blog.native-instruments.com/granular-synthesis/

[9] M. N. Grimshaw, "The Audio Uncanny Valley: Sound, Fear and the Horror Game, " The Audio Uncanny Valley: Sound, Fear and the Horror Game, pp. 21–26, Sep. 2009.

[10] W. De Baene, A. Vandierendonck, M. Leman, A. Widmann, and M. Tervaniemi, "Roughness perception in sounds: behavioral and ERP evidence, " Biological Psychology, vol. 67, no. 3, pp. 319–330, Nov. 2004, doi: https: //doi.org/10.1016/j.biopsycho.2004.01.003.

[11] J. C. Hailstone, R. Omar, S. M. D. Henley, C. Frost, M. G. Kenward, and J. D. Warren, "It's Not What You play, it's How You Play it: Timbre Affects Perception of Emotion in Music, " Quarterly Journal of Experimental Psychology, vol. 62, no. 11, pp. 2141–2155, Nov. 2009, doi: https: //doi.org/10.1080/17470210902765957.

[12] J. Armitage, Imre Lahdelma, Tuomas Eerola, and Rytis Ambrazevičius, "Culture Influences Conscious Appraisal of, but Not Automatic Aversion to, Acoustically Rough Musical Intervals, " PloS One, vol. 18, no. 12, pp. e0294645–e0294645, Dec. 2023, doi: https: //doi.org/10.1371/journal.pone.0294645.

[13] G. A. Bryant and P. E. Smaldino, "The Cultural Evolution of Distortion in Music (and Other Norms of Mixed appeal), " Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 380, no. 1923, Apr. 2025, doi: https: //doi.org/10.1098/rstb.2024.0014.

[14] G. Bloothooft, E. Bringmann, M. van Cappellen, J. B. van Luipen, and K. P. Thomassen, "Acoustics and Perception of Overtone Singing, " The Journal of the Acoustical Society of America, vol. 92, no. 4, pp. 1827–1836, Oct. 1992, doi: https: //doi.org/10.1121/1.403839.

[15] "Sound In A Nutshell: Granular Synthesis, " Granularsynthesis.com, 2025. https: //www.granularsynthesis.com/hthesis/sync.html (accessed Sep. 17, 2025).

[16] S. Riedel, M. Frank, and F. Zotter, "The Effect of Temporal and Directional Density on Listener Envelopment, " Journal of The Audio Engineering Society, vol. 71, no. 7/8, pp. 455–467, Jul. 2023, doi: https: //doi.org/10.17743/jaes.2022.0088.

[17] Eldar Tagi, "Synthesis Methods Explained: What Is Physical Modeling?, " Perfect Circuit, Sep. 19, 2023. https: //www.perfectcircuit.com/signal/what-is-physical-modeling?srsltid=AfmBOopDXiQJn-PSUbqtvdj_Lw_HhhmHRIDYjDD60G0DviD3O3OMPkeb (accessed Sep. 17, 2025).

[18] A. Douglas, "Excite and Resonate: A History of Physical Modelling Synthesis, " Attack Magazine, Feb. 23, 2024. https: //www.attackmagazine.com/features/long-read/excite-and-resonate-a-history-of-physical-modelling-synthesis/

[19] G. Percival, N. Bailey, and G. Tzanetakis, "Physical Modeling Meets Machine learning: Teaching Bow Control to a Virtual Violinist, " ResearchGate, 2011. https: //www.researchgate.net/publication/267236503_Physical_modeling_meets_machine_learning_Teaching_bow_control_to_a_virtual_violinist (accessed Sep. 17, 2025).